

Complete Nucleotide and Deduced Amino Acid Sequence of Rice homolog of MLH1

1	CGGCACGAGATTTTGCAGTCTCCTCTCCTCCTCCGCTCGAGCGAGTGAGTCCCGACCACG	60
61	TCGCTGCCCTCGCCTCACCGCCGGCCAACCGCCGTGACGAGAGATCGAGCAGGGCGGGGC	120
121	ATGGACGAGCCTTCGCCGCGCGGAGGTGGGTGCGCCGGGGAGCCGCCCCGCATCCGGAGG MetAspGluProSerProArgGlyGlyGlyCysAlaGlyGluProProArgIleArgArg	180
181	TTGGAGGAGTCGGTGGTGAACCGCATCGCGGCGGGGGAGGTGATCCAGCGCCGTCGTCG LeuGluGluSerValValAsnArgIleAlaAlaGlyGluValIleGlnArgProSerSer	240
241	GCGGTGAAGGAGCTCATCGAGAACAGCCTCGACGCTGGCGCCTCCAGCGTCTCCGTTGCG AlaValLysGluLeuIleGluAsnSerLeuAspAlaGlyAlaSerSerValSerValAla	300
301	GTGAAGGACGGTGGCCTCAAGCTCATCCAGGTCTCCGATGACGGCCATGGCATCAGGTTT ValLysAspGlyGlyLeuLysLeuIleGlnValSerAspAspGlyHisGlyIleArgPhe	360
361	GAGGATTTGGCAATATTGTGCGAAAGGCATACTACCTCAAAGTTATCTGCATACGAGGAT GluAspLeuAlaIleLeuCysGluArgHisThrThrSerLysLeuSerAlaTyrGluAsp	420
421	CTGCAGACCATAAAATCGATGGGGTTTCAGAGGGGAGGCTTTGGCTAGTATGACTTATGTT LeuGlnThrIleLysSerMetGlyPheArgGlyGluAlaLeuAlaSerMetThrTyrVal	480
481	GGCCATGTTACCGTGACAACGATAACAGAAGGCCAATTGCACGGCTACAGGGTTTCTTAC GlyHisValThrValThrThrIleThrGluGlyGlnLeuHisGlyTyrArgValSerTyr	540
541	AGAGATGGTGTAAATGGAGAATGAGCCTAAGCCTTGCCTGCGGTGAAAGGAACTCAAGTC ArgAspGlyValMetGluAsnGluProLysProCysAlaAlaValLysGlyThrGlnVal	600
601	ATGGTTGAAAATCTATTTTACAACATGGTAGCCCGCAAGAAAACATTGCAGAACTCCAAT MetValGluAsnLeuPheTyrAsnMetValAlaArgLysLysThrLeuGlnAsnSerAsn	660
661	GATGACTACCCCAAGATCGTAGACTTCATCAGTCGGTTTGCAGTCCATCACATCAACGTT AspAspTyrProLysIleValAspPheIleSerArgPheAlaValHisHisIleAsnVal	720
721	ACCTTCTCTTGCAGAAAGCATGGAGCCAATAGAGCAGATGTTTCATAGTGCAAGTACATCC ThrPheSerCysArgLysHisGlyAlaAsnArgAlaAspValHisSerAlaSerThrSer	780
781	TCAAGGTTAGATGCTATCAGGAGTGTCTATGGGGCTTCTGTCTGTCGTGATCTCATAGAA SerArgLeuAspAlaIleArgSerValTyrGlyAlaSerValValArgAspLeuIleGlu	840

FIGURE 1A

841 ATAAAGGTTTCATATGAGGATGCTGCAGATTCAATCTTCAAGATGGATGGTTACATCTCA 900
IleLysValSerTyrGluAspAlaAlaAspSerIlePheLysMetAspGlyTyrIleSer

901 AATGCAAATTATGTGGCAAAGAAGATTACAATGATTCTTTTCATAAATGATAGGCTTGTA 960
AsnAlaAsnTyrValAlaLysLysIleThrMetIleLeuPheIleAsnAspArgLeuVal

961 GACTGTACTGCTTTGAAAAGAGCTATTGAATTTGTGTACTCTGCAACATTGCCTCAAGCA 1020
AspCysThrAlaLeuLysArgAlaIleGluPheValTyrSerAlaThrLeuProGlnAla

1021 TCCAAACCTTTCATATACATGTCCATACATCTTCCATCAGAACACGTGGATGTTAATATA 1080
SerLysProPheIleTyrMetSerIleHisLeuProSerGluHisValAspValAsnIle

1081 CACCCAACCAAGAAAGAGGTTAGCCTTTTGAATCAAGAGCGTATTATTGAAACAATAAGA 1140
HisProThrLysLysGluValSerLeuLeuAsnGlnGluArgIleIleGluThrIleArg

1141 AATGCTATTGAGGAAAACTGATGAATTCTAATACAACCAGGATATTCCAACTCAGGCA 1200
AsnAlaIleGluGluLysLeuMetAsnSerAsnThrThrArgIlePheGlnThrGlnAla

1201 TTAAACTTATCAGGGATTGCTCAAGCTAACCACAAAAGGATAAGGTTTCTGAGGCCAGT 1260
LeuAsnLeuSerGlyIleAlaGlnAlaAsnProGlnLysAspLysValSerGluAlaSer

1261 ATGGGTTCTGGAACAAAATCTCAAAAATTCCTGTGAGCCAAATGGTCAGAACAGATCCA 1320
MetGlySerGlyThrLysSerGlnLysIleProValSerGlnMetValArgThrAspPro

1321 CGCAATCCATCTGGAAGATTGCACACCTACTGGCACGGGCAATCTTCAAATCTTGAAAAG 1380
ArgAsnProSerGlyArgLeuHisThrTyrTrpHisGlyGlnSerSerAsnLeuGluLys

1381 AAATTTGATCTTGTATCTGTAAGAAATGTTGTAAGATCAAGGAGAAACCAAAAAGATGCT 1440
LysPheAspLeuValSerValArgAsnValValArgSerArgArgAsnGlnLysAspAla

1441 GGTGATTTGTCAAGCCGTCATGAGCTCCTTGTGGAAATAGATTCTAGCTTCCATCCTGGC 1500
GlyAspLeuSerSerArgHisGluLeuLeuValGluIleAspSerSerPheHisProGly

1501 CTTTTGGACATTGTCAAGAACTGCACATATGTTGGACTTGCCGATGAAGCCTTTGCTTTG 1560
LeuLeuAspIleValLysAsnCysThrTyrValGlyLeuAlaAspGluAlaPheAlaLeu

1561 ATACAACACAATACCCGCTTATACCTTGTAATGTGGTAAATATTAGTAAAGAACTTATG 1620
IleGlnHisAsnThrArgLeuTyrLeuValAsnValValAsnIleSerLysGluLeuMet

1621 TACCAGCAAGCTTTGTGCCGTTTTGGGAACCTTCAATGCTATTTCAGCTCAGTGAACCAGCT 1680
TyrGlnGlnAlaLeuCysArgPheGlyAsnPheAsnAlaIleGlnLeuSerGluProAla

FIGURE 1B

1681 CCACTTCAGGAGTTGCTGGTGATGGCACTGAAAGACGATGAATTGATGAGTGATGAAAAG 1740
ProLeuGlnGluLeuLeuValMetAlaLeuLysAspAspGluLeuMetSerAspGluLys

1741 GATGATGAGAACTGGAGATTGCAGAAGTAAACACTGAGATACTAAAAGAAAATGCTGAG 1800
AspAspGluLysLeuGluIleAlaGluValAsnThrGluIleLeuLysGluAsnAlaGlu

1801 ATGATTAATGAGTACTTTTCTATTACATTGATCAAGATGGCAAATTGACAAGACTTCCT 1860
MetIleAsnGluTyrPheSerIleHisIleAspGlnAspGlyLysLeuThrArgLeuPro

1861 GTTGTACTGGACCAGTACACCCCTGATATGGACCGTCTTCCAGAATTTGTGTTGGCTTTA 1920
ValValLeuAspGlnTyrThrProAspMetAspArgLeuProGluPheValLeuAlaLeu

1921 GGAAATGATGTTACTTGGGATGACGAGAAAGAGTGCTTCAGAACAGTAGCTTCTGCTGTA 1980
GlyAsnAspValThrTrpAspAspGluLysGluCysPheArgThrValAlaSerAlaVal

1981 GGAACTTCTATGCACTTCATCCCCCAATCCTTCCAAATCCATCTGGGAATGGCATTTCAT 2040
GlyAsnPheTyrAlaLeuHisProProIleLeuProAsnProSerGlyAsnGlyIleHis

2041 TTATACAAGAAAAATAGAGATTCAATGGCTGATGAACATGCTGAGAATGATCTAATATCA 2100
LeuTyrLysLysAsnArgAspSerMetAlaAspGluHisAlaGluAsnAspLeuIleSer

2101 GATGAAAATGACGTTGATCAAGAACTTCTTGCGGAAGCAGAAGCAGCATGGGCCCAACGT 2160
AspGluAsnAspValAspGlnGluLeuLeuAlaGluAlaGluAlaAlaTrpAlaGlnArg

2161 GAGTGGACCATTTCAGCATGTCTTGTGTTCCATCCATGCGACTTTTCCTCAAGCCCCGAAG 2220
GluTrpThrIleGlnHisValLeuPheProSerMetArgLeuPheLeuLysProProLys

2221 TCAATGGCAACAGATGGAACGTTTGTGCAGGTGCTTCCTTGGAGAACTCTACAAGATT 2280
SerMetAlaThrAspGlyThrPheValGlnValAlaSerLeuGluLysLeuTyrLysIle

2281 TTTGAAAGGTGTTAGCTCATAAGTGAGAAAATGAAGGCAGAGTAAGATCATGATTCATGG 2340
PheGluArgCysEnd

2341 AGTGTTTTTGAAAATGTGTATAATTTACCGTATTATGTACTTTGATAGTGTCTGTAGAA 2400

2401 ACTGAAGAAAGAAAGATGGCTTTACTTCTGAATTGAAAGTTAACGATGCCAGCAATTGTA 2460

2461 TATTCTGATCAACCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2501

FIGURE 1C

AminoAcid Sequence of Rice Homolog of MLH1.

1 MDEPSPRGGG CAGEPPRIIR LEESVUNRIA AGEVIQRPSS AVKELIENSL
51 DAGASSVSVA VKDGGLKLIQ VSDDGHGIRF EDLAILCERH TTSKLSAYED
101 LQTIKSMGFR GEALASMTYV GHVTVTTITE GQLHGYRVS Y RDGVMENEPK
151 PCAAVKGTQV MVENLFYNMV ARKKTQLQNSN DDYPKIVDFI SRFVHHINV
201 TFSCRKHGAN RADVHSASTS SRLDAIRSVY GASVVRDLIE IKVSYEDAAD
251 SIFKMDGYIS NANYVAKKIT MILFINDRLV DCTALKRAIE FVYSATLPQA
301 SKPFIYMSIH LPSEHVDVNI HPTKKEVSL NQERIIETIR NAIEEKLMS
351 NTTRIFQTQA LNLSGIAQAN PQKDKVSEAS MSGTKSQKI PVSQMVRTDP
401 RNPSGRLHTY WHGQSSNLEK KFDLVSVRNV VRSRRNQKDA GDLSSRHELL
451 VEIDSSFHPG LLDIVKNCTY VGLADEAFAL IQHNTRLYLV NVVNISKELM
501 YQQALCRFGN FNAIQLSEPA PLQELLMAL KDELMSDEK DDEKLEIAEV
551 NTEILKENAE MINEYFSIHI DQDGKLTRL P VVLDQYTPDM DRLPEFVLAL
601 GNDVTWDEK ECFRTVASAV GNFYALHPPI LPNPSGNGIH LYKKNRDSMA
651 DEHAENDLIS DENDVDQELL AEAEAAWAQR EWTIQHVLFP SMRLFLKPPK
701 SMATDGTFVQ VASLEKLYKI FERC*

mutL/PMS1 signature sequence is shown in bold.

FIGURE 2


```

2 DEPSRGGGCAGEPPRIIRLEESVVNRIAAGEVIQRPSSAVKELIENSLD 51
:| || |||:|.||||| |||||:||||
13 EEESPATTIVPREPPKIQRLEESVVNRIAAGEVIQRPVSAVKELVENSLD 62
52 AGASSVSVAVKDGGLKLIQVSDDGHGIRFEDLAILCERHTTSKLSAYEDL 101
|.||:| ||||| ||||| || ||||| |||. :||
63 ADSSSISVVVKDGGLKLIQVSDDGHGIRREDLPILCERHTTSKLTKFEDL 112
102 QTIKSMGFRGEALASMTYVGHVTVTTITITEGQLHGVRVSYRDGVMEENPKP 151
.: ||||| ||||| |||||:||||| |||||
113 FSLSSMGFRGEALASMTYVAHVTVTTITKGQIHGVRVSYRDGVMEHEPKA 162
152 CAAVKGTQVMVENLFYNMVARKKTLQNSNDYPKIVDFISRFVHHINVT 201
|||||:|||||:|:|||| | || | || |:| |:| ||.
163 CAAVKGTQIMVENLFYNMIARRKTLQNSADDYGKIVDLLSRMAIHNNVS 212
202 FSCRKHGANRADVHSASTSSRLDAIRSVYGASVVRDLIEIKVSYEDAADS 251
||||| :|||| . ||||. |||| | | :. . . : | | ..
213 FSCRKHGAVKADVHSVSPSRLDSIRSVYGVSAKNLMKVEVSSCDSSGC 262
252 IFKMDGYISNANYVAKKITMILFINDRLVDCTALKRAIEFVYSATLPQAS 301
| |:|:| ||. ||||| :||| |||||:|. ||||| ||. |||||
263 TFDMEGFISNSNYVAKKITLVLFINDRLVECSALKRAIEIVYAATLPKAS 312
302 KPFIYMSIHLPSHVVDVNIHPTKKEVSLNQEIRIETIRNAIEEKLMSN 351
|||:| |||. | ||||:||||| ||||| || | .. : | | | . |
313 KPFVYMSINLPREHVDINIHPTKKEVSLNQEIIEMIQSEVEVKLRNAN 362
352 TTRIFQTQALNLSGIAQANPQKDKVSEASMSGTKSQKIPVSQMVRTDPR 401
|| || | . || || || | || | || | || | || | || |
363 DTRTFQEQKVEYIQ. STLTQSQSDSPVSQKPSGQKTQKVPVNKMVRTDSS 411
402 NPSGRLHTYWHGQSSNLEKKFDLVS. VRNVVRSRRNQKDAGDLSSRHELL 450
.|. |||| : . . | | . | ||. || || | : |||| ||:
412 DPAGRLHAFLOPKPQSLPDKVSSLSVVRSSVRQRNPKETADLSSVQELI 461
451 VEIDSSFHPLGLLDIVKNCTYVGLADEAFALIQHNTRLVYNVNNISKELM 500
: || |||:|: |: |||||:|: |||:|: | | ||||: |||||
462 AGVDSCCHPGMLETVRNCTYVGMADDVFALVQYNTHLYLANVVNLSKELM 511
501 YQQALCRFGNFNAIQLSEPAPLQELLVMALKDDEL. .MSDEKDDEKLEIA 548
||| | || . |||||: |||| ||: .: |||: : | . | || | ||
512 YQQTLLRRFAHFNAIQLSDPAPLSELILLALKEEDLDPGNDTKDDLKERIA 561
549 EVNTEILKENAEMINEYFSIHIDQDGKLTLPVVDQYTPDMDRLPFVL 598
|. |||: || |||: ||||: ||| |. ||||: ||||| |||||. |||. |
562 EMNTELLKEKAEMLEEYFSVHIDSSANLSRLPVILDQYTPDMDRVPEFL 611

```

FIGURE 3A

612 CLGNDVEWEDEKSCFQGVSAAGNFYAMHPPLLPNPSGDGIQFYSKRGES 661

712 PASMASNGTFVKVASLEKLYKIFERC 737

Deduced amino acid sequences of *Oryza sativa* and *Arabidopsis thaliana* (Genbank ID, SP_PL:Q9ZRV4) were compared using the Bestfit program of GCG.

FIGURE 3B

FIGURE 4A

[illegible]

FIGURE 4B

1491	CCATCCTGGCCCTTTTGGACATTGTCAAGAACTGCACATATGTTGGACTTG	1540
1406	CCATCCAGGTATGCTGGAGACTGTAAGGAATTGCACATATGTTGGAATGG	1455
1541	CCGATGAAGCCTTTGCTTTGATACAACACAATACCCGCTTATACCTTGTA	1590
1456	CAGATGATGTTTTTGGCTTTAGTTTCAGTATAACACCCATCTATATCTAGCA	1505
1591	AATGTGGTAAATATTAGTAAAGAACTTATGTACCAGCAAGCTTTGTGCCG	1640
1506	AATGTGGTGAATCTCAGCAAAGAGCTAATGTATCAGCAAACCTCTTCGTCG	1555
1641	TTTTGGGAACTTCAATGCTATTCAGCTCAGTGAACCAGCTCCACTTCAGG	1690
1556	TTTTGCTCATTTTAACGCAATACAGCTTAGCGATCCAGCCCCTTTGTCTAG	1605
1691	AGTTGCTGGTGATGGCACTGAAAGACGATGA.ATTGAT.....GAGTGAT	1734
1606	AGTTGATATTGTTGGCTCTGAAAGAGGAGGATCTAGATCCAGGAAATGAT	1655
1735	GAAAAGGATGATGAGAAACTGGAGATTGCAGAAGTAAACACTGAGATACT	1784
1656	ACAAAAGATGATCTGAAAGAAAGAAATTGCTGAAATGAATACAGAACTCCT	1705
1785	AAAAGAAAATGCTGAGATGATTAATGAGTACTTTTTCTATTTCACATTGATC	1834
1706	CAAGGAAAAAGCAGAAATGTTAGAGGAGTATTTTCAGCGTGCACATTGACT	1755
1835	AAGATGGCAAATTGACAAGACTTCCTGTTGTACTGGACCAGTACACCCCT	1884
1756	CCAGTGCAAATTTGTCAAGGCTTCCTGTGATACTCGACCAGTATACACCT	1805
1885	GATATGGACCGTCTTCCAGAATTTGTGTTGGCTTTAGGAAATGATGTTAC	1934
1806	GACATGGATCGTGTTCCTGAATTTTTTACTATGCTTGGGAAATGATGTTGA	1855
1935	TTGGGATGACGAGAAAGAGTGCTTCAGAACAGTAGCTTCTGCTGTAGGAA	1984
1856	GTGGGAAGATGAGAAGAGTTGCTTTCAAGGAGTTTCTGCAGCTATTGGGA	1905
1985	ACTTCTATGCACTTCATCCCCAATCCTTCCAAATCCATCTGGGAATGGC	2034
1906	ACTTTTACGCCATGCATCCTCCTCTTTTGCCAAACCCATCGGGTGACGGT	1955
2035	ATTCATTTATACA.....AGAAAAATAGAGATTC	2063
1956	ATTCAGTTCTATAGTAAGAGAGGTGAGAGCTCTCAGGAAAAGTCAGATTT	2005
2064	AATGGCTGATGAACATGCTGAGAATGATCTAATATCAGATGAAAATGACG	2113
2006	AGAGGGTAACGTCGATATGGAGGACAATC.....	2034
2114	TTGATCAAGAACTTCTTGCGGAAGCAGAAGCAGCATGGGCCCAACGTGAG	2163
2035	TTGACCAAGATCTTCTGTCTAGATGCTGAAAACGCATGGGCACAACGTGAA	2084

FIGURE 4C


```

2164 TGGACCATT CAGCATGTCTTGTTTCCATCCATGCGACTTTTCCTCAAGCC 2213
      ||| | | | | | | | | | | | | | | | | | | | | | | | | |
2085 TGGTCAATCCAACACGTGTTGTTTCCGTCAATGAGATTGTTCTTGAAGCC 2134
      .
2214 CCCGAAGTCAATGGCAACAGATGGAACGTTTGTGCAGGTTGCTTCCTTGG 2263
      || | | | | | | | | | | | | | | | | | | | | | | | |
2135 ACCAGCTTCCATGGCTTCAAATGGGACTTTTGTAAAGGTAGCATCCCTTG 2184
      .
2264 AGAAACTCTACAAGATTTTTTGAAAGGTGTTAGCTCATA 2301
      | | | | | | | | | | | | | | | | | | | | | | | |
2185 AAAAGCTGTACAAGATATTCTGAACGATGCTAACTGAAA 2222

```

[illegible]

FIGURE 4D